

Reproducibility in systems biology modelling

Krishna Tiwari ^{1 2}, Sarubini Kananathan ¹, Matthew G Roberts ¹, Johannes P Meyer ¹,
Mohammad Umer Sharif Shohan ¹, Ashley Xavier ¹, Matthieu Maire ¹, Ahmad Zyoud ¹
Jinghao Men ¹, Szeyi Ng ¹, Tung V N Nguyen ¹, Mihai Glont ¹, Henning Hermjakob * ^{1 3},
Rahuman S. Malik-Sheriff * ¹

1. European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK
2. Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK
3. Beijing Proteome Research Center, Beijing Institute of Lifeomics, National Center for Protein Sciences (The Phoenix Center), Beijing 102206, China

Corresponding authors: hhe@ebi.ac.uk , sheriff@ebi.ac.uk

Abstract

The reproducibility crisis has emerged as an important concern across many fields of science including life science, since many published results failed to reproduce. Systems biology modelling, which involves mathematical representation of biological processes to study complex system behaviour, was expected to be least affected by this crisis. While lack of reproducibility of experimental results and computational analysis could be a repercussion of several compounded factors, it was not fully understood why systems biology models with well-defined mathematical expressions fail to reproduce and how prevalent it is. Hence, we systematically attempted to reproduce **455** kinetic models of biological processes published in peer-reviewed research articles from **152** journals; which is collectively a work of about **1400** scientists from **49** countries. Our investigation revealed that about half (**49%**) of the models could not be reproduced using the information provided in the published manuscripts. With further effort, an additional 12% of the models could be reproduced either by empirical correction or support from authors. The other 37% remained non-reproducible models due to missing parameter values, missing initial concentration, inconsistent model structure, or a combination of these factors. Among the corresponding authors of the non-reproducible model we contacted, less than **30%** responded. Our analysis revealed that models published in journals across several fields of life science failed to reproduce, revealing a common problem in the peer-review process. Hence, we propose an 8-point reproducibility scorecard that can be used by authors, reviewers and journal editors to assess each model and address the reproducibility crisis.

Tiwari *et al.*

Introduction

Reproducibility of scientific results is a key determinant of true science and credibility. The reproducibility crisis across many fields of science has emerged as an important concern when results published by leading scientific journals failed to reproduce¹⁻⁴. In a survey of 1,576 scientists published in Nature⁵, about 90% acknowledged this emerging crisis. The survey reported that over 70% scientists failed to reproduce others' experiments and over 50% failed to reproduce their own results. Experimental results fail reproducibility tests due to several reasons including improper documentation of methodology, considering noise as a positive finding, unrecognized or incomplete experimental variables, data fabrication or bias, publishing premature or incomplete results^{6-9,5} and inappropriate statistical analysis (e.g. "P-hacking")^{10,11}. Computational biology research also faces this crisis, and the lack of reproducibility is compounded by several factors including changes in reference data and/or formats, software versions, missing essential codes or methodology etc.¹²⁻¹⁴. Several suggestions have been published to improve reproducibility in computational biology^{12,15,16} and bioinformatics^{17,18}.

Systems biology modelling involves mathematical representation of biological processes to investigate complex behaviour of the system which cannot be studied by looking at individual components¹⁹⁻²¹. It was supposed that modelling will remain relatively untouched by this crisis, as the models are a specific set of computational codes representing well defined mathematical equations to perform reproducible simulations. However, mathematical models from a number of manuscripts do not reproduce the simulation results²²⁻²⁴ described in the manuscript. While lack of reproducibility of experiment results and computational analysis could be a repercussion of several compounded factors, lack of reproducibility in the simulation of mathematical equations is typically due to inadvertent error or lack of information in the manuscript. It is critical to recognise the cause of models' failure to surmount the reproducibility crisis. Moreover, it was also not clear how prevalent the reproducibility crisis is in the field of systems biology modelling.

Hence, to investigate the reproducibility crisis, we systematically analyzed mathematical models in conjunction with the curation process in BioModels repository²⁵. BioModels (<https://www.ebi.ac.uk/biomodels/>) is one of the largest public open source databases of quantitative mathematical models, where the models are manually curated and semantically enriched. In total we investigated 455 kinetic models of biological processes published in literature; which is collectively the work of about 1400 scientists from 49 countries. It is one of the largest studies performed to assess the status of reproducibility by attempting to independently reproduce results published in a number of research articles. We could reproduce about half of the models (51%), whereas the other half could not be reproduced using the information provided in the manuscript. We managed to reproduce a further 9% of the models with empirical trial and error approach and an additional 3% with the help of authors who responded to our email requests. Our in-depth investigation has revealed the major reasons why models fail to reproduce the published simulation results. We discuss these reasons and further propose a reproducibility scorecard for modellers, reviewers and journal editors to assess models and address the reproducibility crisis.

Methodology

To investigate the reproducibility in the systems biology models, we systematically analyzed 455 ordinary differential equation (ODE) models from published research articles. This exercise was done in conjunction with the curation of models in BioModels. The manual curation of models in BioModels involves a two-step process (1) encoding models in standard formats and reproducing the simulation figures in the reference manuscript and (2) semantic enrichment of the model and its components²⁵. The reproducibility assessment was done along with the first step of the model curation. Semantic enrichment of the model in BioModels was done following MIRIAM guidelines²⁶ and it involved annotation of model entities (species, reactions, parameters, events, etc) with cross-references to controlled vocabularies such as GO^{27,28}, ChEBI²⁹, Mathematical Modelling Ontology³⁰, Systems Biology Ontology³¹, Brenda Tissue Ontology³² and Experimental Factor Ontology³³, as well as data resources such as UniProt³⁴, Ensembl³⁵, NCBI Taxonomy³⁶, Reactome³⁷, etc. The models for curation are selected either from those submitted to BioModels by modellers or from the published articles contingent on the interest of the curators and collaborators. Our collection covered models of a wide range of biological processes, originating from articles published in 152 life science journals.

Reproducibility assessment

The following steps were employed in our study to assess the reproducibility:

1. The model manuscript was carefully read and the model equations were encoded in the standard SBML³⁸ format. When the models were previously submitted in SBML format, the equations, values of parameters and initial concentration, perturbation events, etc. were cross-verified with the reference manuscript.
2. The simulation of SBML model files was performed predominantly using COPASI³⁹. When COPASI was used in the original manuscript, other simulation software such as simbiology toolbox (MATLAB)⁴⁰, libSBMLsim⁴¹, Mathematica⁴² were used to perform simulations.
3. The model was considered as reproducible when it reproduced at least one of the main simulation figures in the associated research article using a software different from the one used in the original manuscript. The reproduced simulation figure, such as time course plot with and without perturbation, phase-plane plot, etc., should precisely match the original figure, any minor deviation was still considered acceptable if it did not affect the scientific conclusion of the study. The models which could be directly reproduced with the description in the manuscript were labelled as 'Directly Reproducible'.
4. When the model failed to reproduce the simulation with the mathematical equation and the parameter values provided in the research article, we resorted to an empirical trial and error approach to correct the model based on curator expertise. For example, any terms missing in the equations but described in the manuscript were added to correct the model; any potential typos such as misplacement of decimal points in the parameter values were corrected. The models that were reproduced after such corrections were labelled 'Reproduced with manual corrections'.

Tiwari *et al.*

5. The authors of the failed, yet potentially salvageable models were contacted when possible and their responses were recorded. The models that were reproducible with the corrections provided by them were labelled as 'Reproduced with author support'.
6. Nevertheless, when the model still could not be reproduced, they were labelled as 'Non-reproducible' and the likely reasons were recorded. The plausible reasons for non-reproducibility include (a) inconsistency in model structure i.e. any error in the model equation (b) missing parameter values (c) missing initial concentration and (d) unknown reason.
7. SBML representation of all the analysed models were submitted to BioModels²⁵. The reproducible ones were labelled as curated models.

Results

About half of the published models were not directly reproducible

In total 455 kinetic models from peer-reviewed and published scientific publications were subjected to reproducibility tests and extensively analysed to unveil the fundamental cause behind systems biology model's failure to reproduce the simulation figures in the associated manuscripts. We selected kinetic model manuscripts published from 1980 to 2020 (Sup Fig 1). Among these models, the mathematical equations of 389 models were manually encoded in the standard SBML format using COPASI³⁹ from the original manuscript. The remaining 66 models were those submitted to BioModels in SBML format by the authors and hence they were carefully cross-checked to ensure whether the mathematical equations, initial conditions and parameters were accurately represented. The SBML representation of 233 out of 455 models (51%) directly reproduced simulation results in the original manuscript (Figure 1a). About 49% of the published models were not reproducible either due to incorrect or missing information in the manuscript. This high proportion is unexpected and exposed a serious issue within the field. Even among the 66 models submitted to BioModels in SBML standard format, only 37 could be reproduced directly.

About 12% of the models could be reproduced with further effort

About 12% of all models could be rescued with further efforts involving either a careful empirical trial and error approach or author support (Figure 1b). 40 models (9%) were successfully reproduced with manual empirical correction of the inaccurate reporting in the manuscript. Some of the common errors that could be identified and manually corrected were (a) error in the sign of the terms in the mathematical equations, e.g. a negative sign for a production term in the equation or vice versa, could be easily identified by expert curation and corrected (b) missing terms in the model equations - e.g. missing one of the production or depletion terms in the ODE definition (c) typos in the parameter values e.g. mistakes in the decimal points, a value of 0.01 was reported in place of 0.001 (d) missing values e.g. some missing initial concentrations of model entities could be inferred from the initial time point in the simulation plots (e) error in the units of initial concentration and parameter values e.g. nM/l was misrepresented as uM/l.

Reproducibility in systems biology modelling

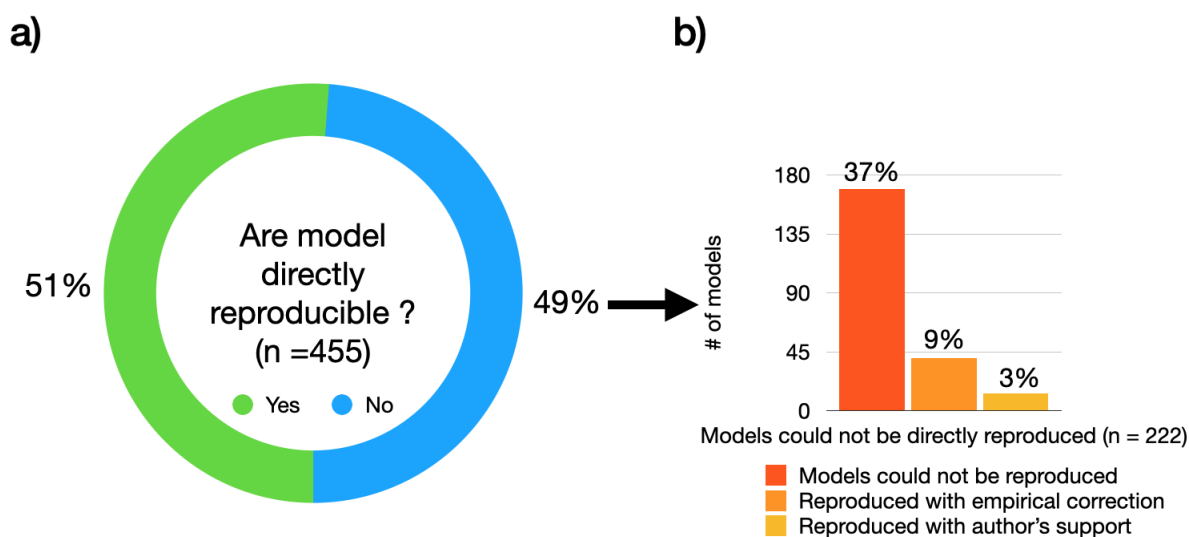


Figure 1: Reproducibility of systems biology models. (a) About half of the published systems biology models could not be directly reproduced. (b) About 12% of the models could be reproduced with empirical corrections or author support.

It was not always possible to estimate the missing concentration, parameter or the mathematical expression. Hence, we contacted corresponding authors to request the missing information or seek clarification. This was not always feasible, for example due to authors' change of institutions, change of field, leaving academia, and death. In total, we attempted to contact corresponding authors of 90 models, among which less than a third (27) responded (Figure 2). About half of the models of authors who responded (13 models, or 3% of the total

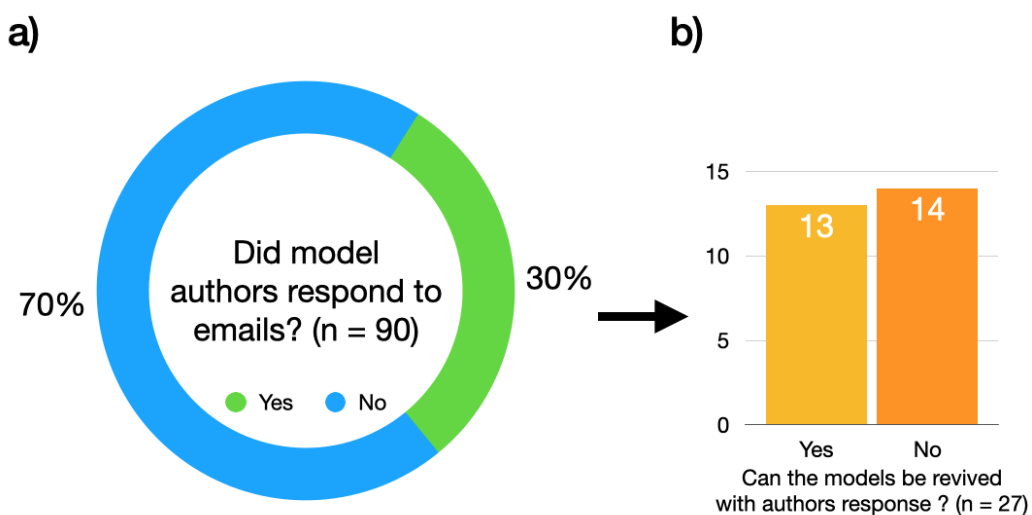


Figure 2: Rescuing reproducibility with author support. (a) A majority of authors of non-reproducible models did not respond to emails. (b) About half of the models of the authors who responded could eventually be reproduced.

Tiwari *et al.*

number of models) were subsequently reproduced. These were mostly models published in the last 5 years (Sup Fig 1). Surprisingly about 70% of the authors we contacted did not respond to the request to provide information to reproduce their models (Figure 2).

The major reasons why models failed to reproduce

About 37% of the 455 models could not be rescued even with further efforts. The major reasons why these models failed to reproduce include missing parameters values as the main reason, followed by missing initial concentration and inconsistency in model structure, or a combination of aforementioned causes (Figure 3). Yet, in a large proportion of the non-reproducible models, the reason for failure was unclear. The reference manuscripts of those models might have probably reported incorrect parameter values, initial concentrations or model equations or a combination of these three factors. Some research articles report parameter values in the form of plots, and it was not straightforward to extract those values to reproduce the simulations. Insufficient and incorrect reporting of the model content were the main reasons why models failed to reproduce. These factors are commonly overlooked in the peer-review process and hence are reflected across several journals where systems biology models are published (Figure 4).

Reason for Non-reproducibility (n = 169)

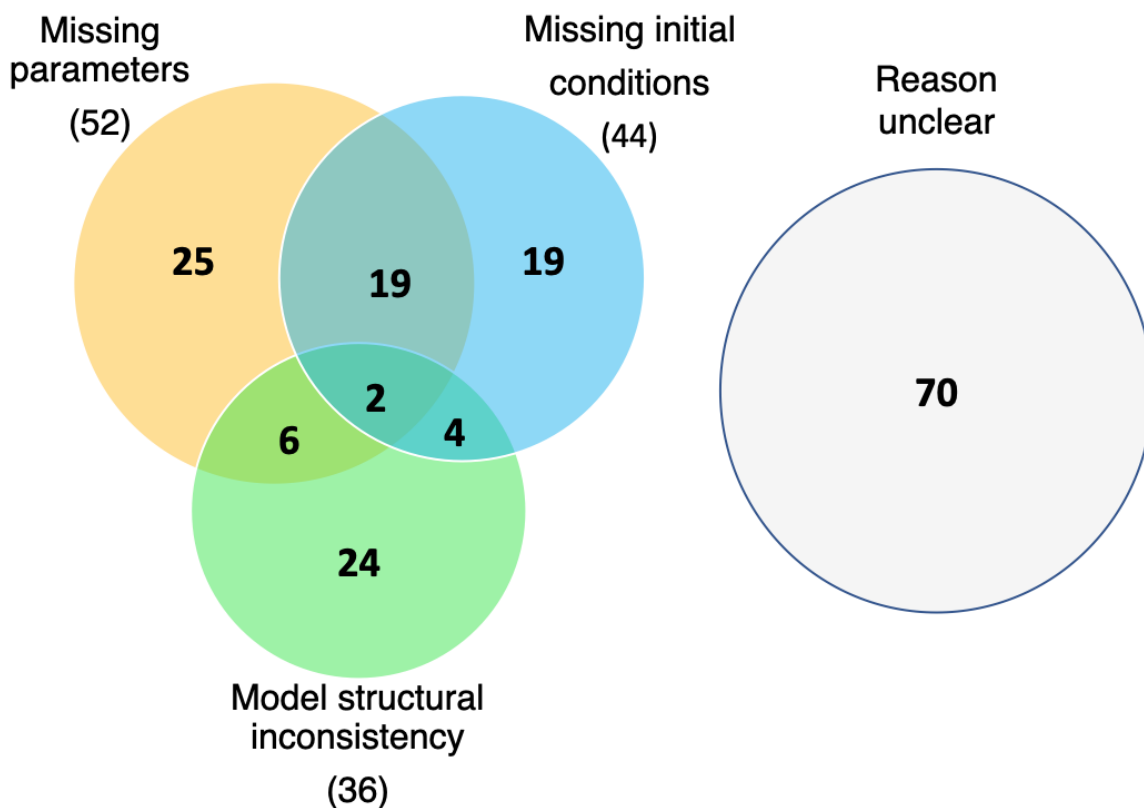


Figure 3: Common reasons why models could not be reproduced include missing parameter values, initial conditions and inconsistency in model structure.

Reproducibility in systems biology modelling

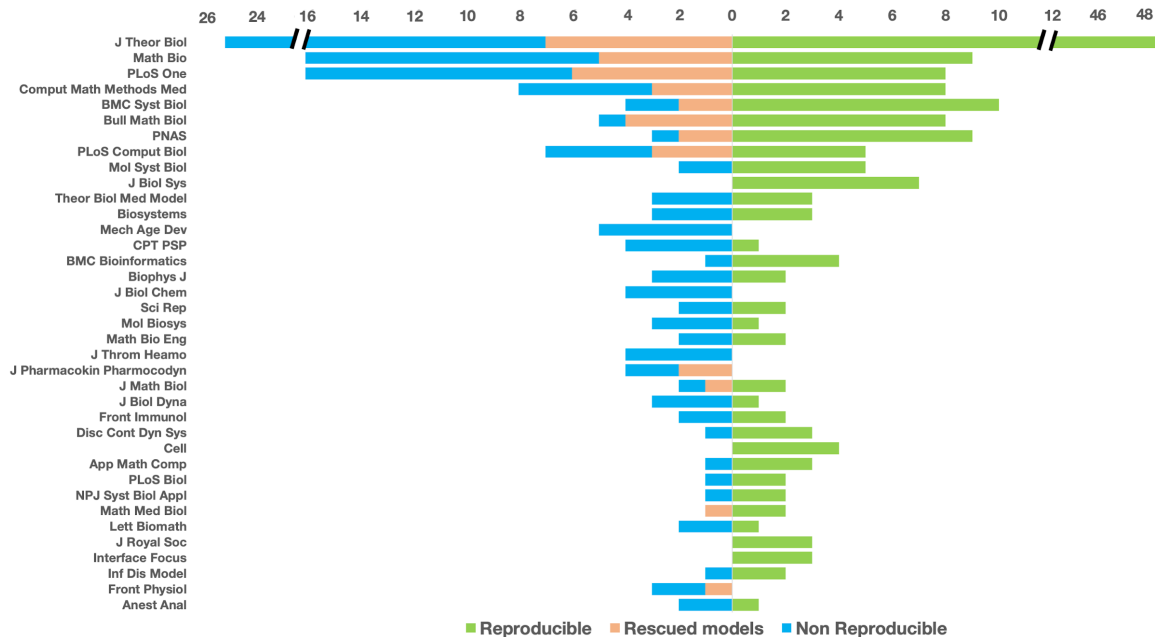


Figure 4: Distribution of reproducible, rescued and non-reproducible models across various journals with more than two models in our study.

Discussion

Reproducibility is the essence of science and any scientific work that cannot be reproduced independently using the same resources is deemed untrustworthy^{22,43}. The ability to reproduce a biological experiment is often limited by several confounding factors and is not unheard of. Similarly, reproducible bioinformatics analysis is affected by regular updates, releases in data and software, etc.¹⁸. Systems biology models are least expected to be affected by the lack of reproducibility, as they are merely a set of mathematical functions that are straight forward to simulate. In contrast, our analysis of 455 ODE models has exposed that about half of these models cannot be reproduced using the information provided in the manuscript. Although the lack of reproducibility has been discussed within the community^{17,12,22}, it was never expected to be this adverse. Our study has revealed that the inability to reproduce models is widespread in research articles published across several journals in the field with the exception of a few (Figure 4), and it is imperative to revisit the peer-review process of mathematical studies as previously suggested²³.

Reproducibility, replicability, repeatability are the terminologies often confused and are defined differently in experimental^{43,5} and computational research⁴⁴. In the context of systems biology modelling, the refined definition of replicability (also referred as repeatability) is the ability to use the same code provided with the manuscript in the same software to reproduce the simulation results; whereas reproducibility is the ability to build the code *de novo* and/or ensure the mathematical expressions are correctly represented and reproduce the simulation results in a software different from the one originally used. The focus of this work was to assess the reproducibility of the published systems biology models and hence the latter was chosen as the criterion.

Tiwari *et al.*

Even in the models that are reproduced, one of the challenges we faced is unambiguous inference of the model entities and their values. When a variable or model entity name is different in the main manuscript description, mathematical expression, and code, it becomes challenging to match them to reproduce the simulation. For example, “alpha” in the model description and/or equation in the manuscript and the code may refer to completely different entities. We have overcome this challenge by carefully reading the reference manuscript. We strongly recommend making the code or the model file as self-contained as possible with proper annotation of the model entities. Model codes written in programming languages such as Matlab, python, C, R, etc. are often helpful to reproduce the model. Nevertheless, not all such codes are easily comprehensible, specifically when they are not well commented. Although the systems modelling community is split, a notable fraction of the modellers use COMBINE⁴⁵ community standard formats such as SBML, SED-ML⁴⁶, COMBINE Archive to encode their models. These standard formats provide a consistent framework to encode and annotate models, making them both human and machine readable. The strong community support for standard formats such as SBML makes it highly interoperable with about 280 supporting software tools for model construction, simulation, visualization and processing the semantic layer. We highly recommend using standard formats to encode and disseminate mathematical models as these greatly enhance the ability to comprehend and reproduce the models.

The most common approaches in systems biology modelling include kinetic, constraint-based, logic and agent-based modelling²¹. In this work, we specifically focused on ODE models, one of the types of kinetic models, to analyse why they cannot be reproduced. Half of these relatively simple deterministic kinetic models could not be reproduced. Other types of kinetic models include delay differential equations and partial differential equations; they are likely to be affected either to the same extent or even more by the reproducibility crisis as they are relatively more complex than ODE models. In the case of constraint-based modelling, flux values resulting from flux balance analysis are commonly reported in manuscripts and these values are not unique solutions and cannot be directly reproduced. The community tool MEMOTE⁴⁷ was developed primarily to quality control constraint based models. We are currently collaborating with the constraint-based modellers to develop tools and procedures to test reproducibility of these models. Similarly, we have engaged with the logic modelling community to develop guidelines for curation and annotation of logic models (CALM)⁴⁸.

Leveraging on the lessons learned from attempting to reproduce 455 models as well as our interaction with modelling communities, we have developed a reproducibility scorecard to enhance the ability to reproduce systems biology models (Box 1). The scorecard consists of a list of items that would allow another modeller to reproduce the simulation results of a model with a reasonable effort. We recommend authors, reviewers and editors of the journals to assess each systems biology model in the research article using our proposed scorecard. The scorecard consists of 8 questions with a unit score for each ‘yes’ as an answer. All the 8 questions may not always be applicable and hence, in the scale of 8, we strongly advocate that a model get a score of 4 at the least.

A clear and complete description of the mathematical model, relevant parameter values and simulation conditions are vital to reproduce the model and are addressed in the first three questions. In case of large models, it may not be possible to enumerate the aforementioned information in the manuscript, hence it is acceptable to share the code publicly either via

Reproducibility in systems biology modelling

Box1	
Reproducibility score card	
1) Are the mathematical expressions described in the manuscript / supplementary material?	
2) Are the parameters and entity initial levels listed as a table in the manuscript / supplementary material?	
3) Are simulation conditions including changes in parameters / concentration / states and any data normalization described under each simulation figure or attached as a supplementary material?	
4) Are the model code(s) for the mathematical expression and simulation shared publicly?	
5) Are the model codes available in standard format such as SBML, COMBINE archive, SED-ML and are syntactically validated?	
6) Are the model codes deposited in a relevant open model database?	
7) Are the model codes well documented to unambiguously identify model entities / variables? (with additional annotation of reactions, mathematical expressions, events, conditions, etc. when relevant.) Are the models in standard formats such as SBML, COMBINE Archive are semantically enriched i.e. annotated with controlled vocabularies such as Gene Ontology, ChEBI, etc and database resources such as Gene Ontologies?	
8) Are the numerical results shared publicly along with the model codes?	
Total Score (out of 8)	

supplementary material, git-hub, or a model repository. Following the Findable, Accessible, Interoperable and Reusable (FAIR) ⁴⁹ principles, we have recommended submission of the model code in standard formats into the relevant model repositories in question 5 and 6 respectively. However, question 4 is not redundant, as it is used to cover those modellers who use non-standard formats, for example as collection of programming scripts and share them via git-hub, unstructured repositories, or similar. A model shared in standard format such as SBML, CellML ⁵⁰ gets an additional score in our scorecard, as the interoperability of the model file will provide the possibility to test and reproduce it using several supporting software tools.

Similarly, deposition of models to open repositories including, but not limited to BioModels, Physiome ⁵⁰ or JWSONline ⁵¹ will get an additional score as they promote FAIR sharing. The advantage of submitting models to open model repositories include provision of (1) a sophisticated search engine to make models findable, (2) a version-controlled storage system to make the models readily accessible, (3) support for interoperable standard formats and (4)

Tiwari *et al.*

curation and annotation services to promote reusability. BioModels, being the largest repository of open, curated, well annotated, and findable models, contributes significantly to reproducibility²². Several journals recommend to submit their model to BioModels⁵². Similar to the curation service in BioModels, JWSONline and the Physiome repository through the Centre for Reproducible BioMedical modelling¹² provide expert curation services to validate published reports.

Unambiguous identification of the model entities is critical to reproduce the results, hence semantic enrichment or proper documentation of the code brings added value. Even an accurately defined model cannot be reproduced if the data normalization and the simulation conditions which include changes to specific parameters or concentration of model entity are not clearly described. Hence, we recommend provision of this information under each simulation figure. Alternatively, the modellers can submit SED-ML files, a COMBINE standard for description of simulation experiments. Although it is possible to use the simulation figures in the manuscript as reference to test the reproducibility of the models in many cases, it is desirable to provide the numerical output of the simulations to verify the model reproducibility.

Model curation is a time intensive task; on average it took about a week to carefully encode and thoroughly investigate the reproducibility of a model; in some cases, it took less than two days and in some cases over two weeks. Our criterion for a reproducible model was that it should reproduce at least one figure from the original article, which we consider to be a reasonable compromise between reliably assessing reproducibility, and the huge additional effort that would be required to ensure reproducibility of all (relevant) figures.

Our study was not intended to call-out non-reproducible models, but rather to highlight the current status of and reasons behind the lack of reproducibility. Thereby, we intend to raise community awareness among the researchers who use systems biology models and provide a potential solution to address the reproducibility crisis. We provide the “curated” section of BioModels as a source of reliable, verified reproducible models, while still keeping the non-reproducible models accessible. The versioning system in BioModels will allow authors to improve their models, while keeping the original version accessible as a part of the public record. Currently, we just keep the non-reproducible models in the “non-curated” part of BioModels, which also contains models from direct submissions which are still awaiting curation, as well as models in representations for which we currently don’t provide detailed curation. We do not explicitly label non-reproducible models because, on the one hand, there is a chance that the failure to reproduce the model is due to curator error, and on the other hand we don’t want to discourage authors from making their models accessible through a public repository. However, we are aware that the lack of explicit labelling of a non-reproducible model might cause others to try the same again. We are open to community suggestions for a more transparent labelling of models.

Lack of reproducibility has an adverse effect on the reliability of the scientific results. An additional consequence is the loss of the time and resources spent by modellers across the globe attempting to reproduce a previously published model. Availability of model code can reduce the time needed to reproduce the model. Nevertheless, submission of model codes in standard format to open repositories is not alone sufficient to reproduce. Among the models we investigated were 66 SBML models submitted by modellers to BioModels, yet nearly half of them (29) could not be reproduced directly either. Hence, we formulated the 8-points

reproducibility scorecard with a 4-point cut-off. BioModels will actively seek to curate models published with our proposed score card. By adapting our scorecard in the peer-review process, complemented by the curation services provided by model repositories and reproducibility centres, we believe that the reproducibility crisis can be significantly addressed. It can be achieved as a community, where authors, reviewers, journal editors embrace reproducibility more proactively than before.

References

1. Begley C. Glenn & Ioannidis John P.A. Reproducibility in Science. *Circ. Res.* **116**, 116–126 (2015).
2. Chen, X. *et al.* Open is not enough. *Nat. Phys.* **15**, 113–119 (2019).
3. Fanelli, D. Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proc. Natl. Acad. Sci.* **115**, 2628–2631 (2018).
4. Munafò, M. R. *et al.* A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 1–9 (2017).
5. Baker, M. 1,500 scientists lift the lid on reproducibility. *Nat. News* **533**, 452 (2016).
6. Puzstai, L., Hatzis, C. & Andre, F. Reproducibility of research and preclinical validation: problems and solutions. *Nat. Rev. Clin. Oncol.* **10**, 720–724 (2013).
7. Errington, T. M. *et al.* An open investigation of the reproducibility of cancer biology research. *eLife* **3**, e04333 (2014).
8. Eisner, D. A. Reproducibility of science: Fraud, impact factors and carelessness. *J. Mol. Cell. Cardiol.* **114**, 364–368 (2018).
9. Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J. & Reinero, D. A. Contextual sensitivity in scientific reproducibility. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6454–6459 (2016).
10. Colquhoun, D. The reproducibility of research and the misinterpretation of p-values. *R. Soc. Open Sci.* **4**, 171085 (2017).
11. Redish, A. D., Kummerfeld, E., Morris, R. L. & Love, A. C. Opinion: Reproducibility failures are essential to scientific inquiry. *Proc. Natl. Acad. Sci.* **115**, 5042–5046 (2018).

Tiwari *et al.*

12. Papin, J. A., Gabhann, F. M., Sauro, H. M., Nickerson, D. & Rampadarath, A. Improving reproducibility in computational biology research. *PLoS Comput. Biol.* **16**, e1007881 (2020).
13. Schaduangrat, N. *et al.* Towards reproducible computational drug discovery. *J. Cheminformatics* **12**, 9 (2020).
14. Stodden, V., Seiler, J. & Ma, Z. An empirical analysis of journal policy effectiveness for computational reproducibility. *Proc. Natl. Acad. Sci.* **115**, 2584–2589 (2018).
15. Leek, J. T. & Peng, R. D. Opinion: Reproducible research can still be wrong: Adopting a prevention approach. *Proc. Natl. Acad. Sci.* **112**, 1645–1646 (2015).
16. Sandve, G. K., Nekrutenko, A., Taylor, J. & Hovig, E. Ten Simple Rules for Reproducible Computational Research. *PLoS Comput. Biol.* **9**, e1003285 (2013).
17. Kim, Y.-M., Poline, J.-B. & Dumas, G. Experimenting with reproducibility: a case study of robustness in bioinformatics. *GigaScience* **7**, (2018).
18. Kulkarni, N. *et al.* Reproducible bioinformatics project: a community for reproducible bioinformatics analysis pipelines. *BMC Bioinformatics* **19**, 349 (2018).
19. Le Novère, N. Quantitative and logic modelling of gene and molecular networks. *Nat. Rev. Genet.* **16**, 146–158 (2015).
20. Fischer, H. P. Mathematical Modeling of Complex Biological Systems. *Alcohol Res. Health* **31**, 49–59 (2008).
21. Lavrik, I. N. & Zhivotovsky, B. Systems biology: a way to make complex problems more understandable. *Cell Death Dis.* **5**, e1256–e1256 (2014).
22. Mendes, P. Reproducible Research Using Biomodels. *Bull. Math. Biol.* **80**, 3081–3087 (2018).
23. Schnell, S. “Reproducible” research in mathematical sciences requires changes in our peer review culture and modernization of our current publication approach. *Bull. Math. Biol.* **80**, 3095–3105 (2018).

24. Crook, S. Model Reproducibility: Overview. in *Encyclopedia of Computational Neuroscience* (eds. Jaeger, D. & Jung, R.) 1–3 (Springer, 2013). doi:10.1007/978-1-4614-7320-6_763-1.
25. Malik-Sheriff, R. S. *et al.* BioModels-15 years of sharing computational models in life science. *Nucleic Acids Res.* **48**, D407–D415 (2020).
26. Le Novère, N. *et al.* Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* **23**, 1509–1515 (2005).
27. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
28. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
29. Hastings, J. *et al.* ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214-1219 (2016).
30. Mathematical Modelling Ontology - Summary | NCBO BioPortal.
<https://bioportal.bioontology.org/ontologies/MAMO>.
31. Courtot, M. *et al.* Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.* **7**, 543 (2011).
32. Gremse, M. *et al.* The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. *Nucleic Acids Res.* **39**, D507–D513 (2011).
33. Malone, J. *et al.* Modeling sample variables with an Experimental Factor Ontology. *Bioinforma. Oxf. Engl.* **26**, 1112–1118 (2010).
34. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
35. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
36. Federhen, S. The NCBI Taxonomy database. *Nucleic Acids Res.* **40**, D136-143 (2012).
37. Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).

Tiwari *et al.*

38. Hucka, M. *et al.* The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 2 Core Release 2. *J. Integr. Bioinforma.* **16**, (2019).
39. Hoops, S. *et al.* COPASI—a COmplex PATHway Simulator. *Bioinforma. Oxf. Engl.* **22**, 3067–3074 (2006).
40. MATLAB. *version 7.10.0 (R2010a)*. (The MathWorks Inc., 2010).
41. Takizawa, H. *et al.* LibSBMLSim: a reference implementation of fully functional SBML simulator. *Bioinformatics* **29**, 1474–1476 (2013).
42. *Mathematica*. (Wolfram Research, Inc., 2019).
43. Goodman, S. N., Fanelli, D. & Ioannidis, J. P. A. What does research reproducibility mean? *Sci. Transl. Med.* **8**, 341ps12-341ps12 (2016).
44. Miłkowski, M., Hensel, W. M. & Hohol, M. Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail. *J. Comput. Neurosci.* **45**, 163–172 (2018).
45. Waltemath, D. *et al.* The first 10 years of the international coordination network for standards in systems and synthetic biology (COMBINE). *J. Integr. Bioinforma.* (2020) doi:10.1515/jib-2020-0005.
46. Waltemath, D. *et al.* Reproducible computational biology experiments with SED-ML—the Simulation Experiment Description Markup Language. *BMC Syst. Biol.* **5**, 198 (2011).
47. Lieven, C. *et al.* MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* **38**, 272–276 (2020).
48. Niarakis, A. *et al.* Setting the basis of best practices and standards for curation and annotation of logical models in biology—highlights of the [BC]2 2019 CoLoMoTo/SysMod Workshop. *Brief. Bioinform.* (2020) doi:10.1093/bib/bbaa046.
49. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, (2016).
50. Lloyd, C. M., Lawson, J. R., Hunter, P. J. & Nielsen, P. F. The CellML Model Repository. *Bioinformatics* **24**, 2122–2123 (2008).

Reproducibility in systems biology modelling

51. Peters, M., Eicher, J. J., van Niekerk, D. D., Waltemath, D. & Snoep, J. L. The JWS online simulation database. *Bioinforma. Oxf. Engl.* **33**, 1589–1590 (2017).
52. Juty, N. *et al.* BioModels: Content, Features, Functionality, and Use. *CPT Pharmacomet. Syst. Pharmacol.* **4**, e3 (2015).